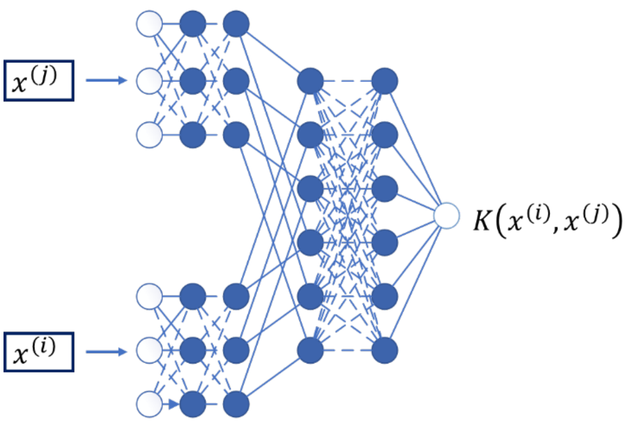
KENNESAW STATE UNIVERSITY

Analytics and Data Science Institute

datascience.kennesaw.edu

Data Science Research Series

Deep Embedding Kernel

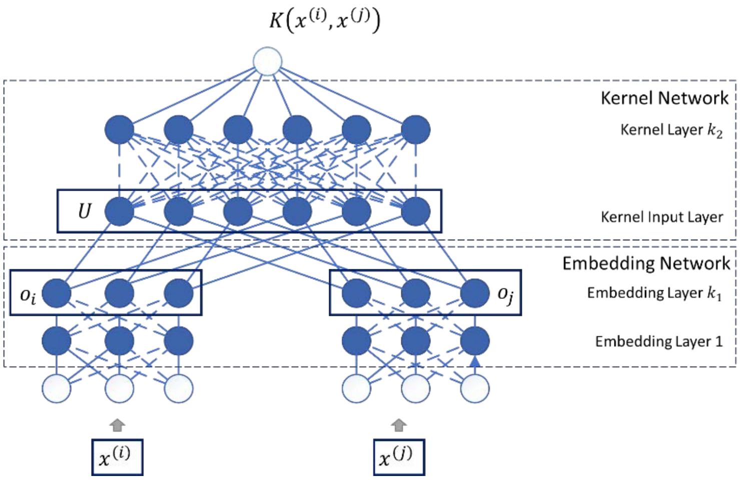


BACKGROUND

This paper proposes a novel supervised learning method that is called Deep Embedding Kernel (DEK). DEK combines the advantages of deep learning and kernel methods in a unified framework. In details, DEK does not explicitly map data to a feature space with pre-specified dimensionality, nor implicitly map data through a pre-defined kernel; instead, DEK uses a newly designed deep architecture to represent a learnable kernel. In other words, DEK utilizes the learning power of deep learning to train a kernel, which in turn implicitly maps data to a high dimensional feature space. The learning objective of DEK specifies a desired relationship of data in the mapped feature space. Therefore, the whole mapped feature space, including its dimensionality, is learned via deep learning. Using deep architectures to learn a kernel, instead of directly learn the feature space also has the advantages of flexibility in that the learned kernel can be applied to a wide range of supervised learning tasks including identity detection, general classification, dimension reduction, regression, and other kernel based machine learning applications. Experiments show that DEK has superior performance over other typical supervised learning methods, such as Kernel Support Vector Machines (SVM), Gradient Boosting Trees (GB), Random Forests (RF), and Neural Networks (MLP) on mentioned learning tasks.

APPROACH

DEK acts as a kernel function: takes two data points as inputs and outputs the kernel value that represents the points' similarity. The DEK architecture consists of a deep embedding network and a deep kernel network, as shown in the figure below:



The embedding network learns an optimized data representation to feed into the kernel network which models the desired data relationship on top of the learned feature space. Both networks are trained simultaneously in a single gradient descent process.

Let o_i and o_j be the representations of the two data points $x^{(i)}$ and $x^{(j)}$ mapped by the embedding network, and U be the input of the kernel network, then

$$U = \begin{Bmatrix} o_{i1} * o_{j1}, o_{i2} * o_{j2}, \dots o_{id} * o_{jd} \\ |o_{i1} - o_{j1}|, \dots |o_{id} - o_{jd}| \end{Bmatrix}$$

The output of DEK directly reflects the similarity (kernel value) of the two data points $x^{(i)}$ and $x^{(j)}$, which is the probability of them having the same labels:

$$K(x^{(i)}, x^{(j)}) = P(y^{(i)} = y^{(j)} \mid x^{(i)}, x^{(j)})$$

DEK can be adjusted to solve various supervised learning tasks including classification, regression, identity detection, and (supervised) dimension reduction. DEK can also work on top of other deep architectures like Convolutional Neural Network or Recurrent Neural Network to model unstructured data (e.g. image, audio, text, etc.)

RESULTS

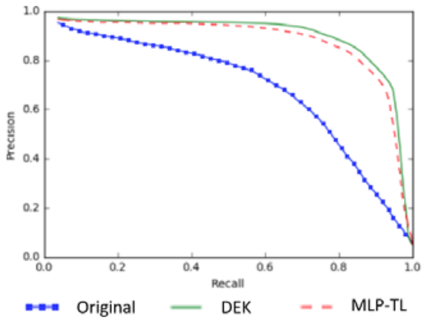
In **classification**, five experiments are conducted, where DEK outperforms all typical classification models. Test accuracy of tested models are shown in the table below:

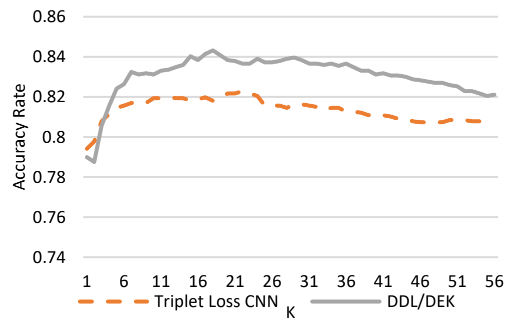
Dataset	SVM/DEK	KNN/DEK	SVM/RBF	GB	RF	MLP
Segment	**0.9691**	0.9678	0.9647	0.9604	0.9610	0.9593
Cardiotocography	0.9893	**0.9899**	0.9879	0.9825	0.9846	0.9850
Messidor Features	**0.7803**	0.7746	0.7543	0.7110	0.7168	0.7222
Waveform	0.8696	**0.8704**	0.8684	0.8488	0.8456	0.8672
Pima Diabete	0.7839	**0.7865**	0.7708	0.7396	0.7604	0.7630

In **regression**, DEK outperforms all models in two over three experiments, and is slightly behind GB regression in one. Test R^2 of tested models are shown in the table below:

Dataset	SVR/DEK	KNN/DEK	SVR/RBF	GB	RF	MLP
Concrete	0.8651	0.8980	0.8702	**0.9067**	0.8751	0.8119
Airfoil	0.8242	**0.9195**	0.8371	0.8840	0.9047	0.8568
Energy Efficiency	0.9685	**0.9783**	0.9621	0.9775	0.9756	0.9470

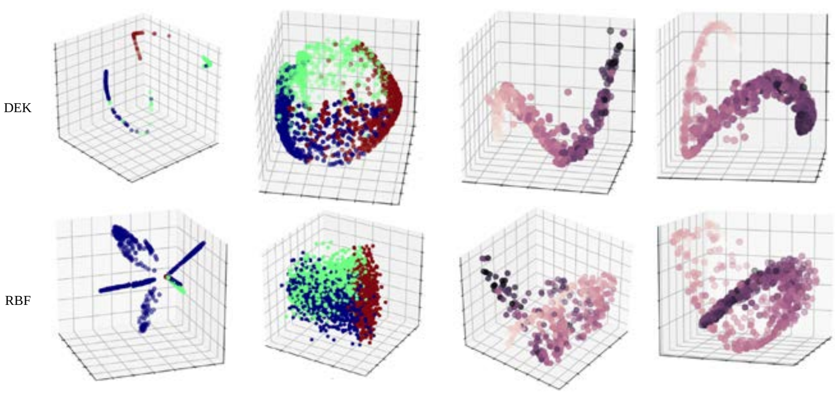
In **identity detection**, two experiments are done, one in **facial recognition** and on in **speaker identification**. In the facial recognition task, DEK is laid on top of a pre-trained Google Facenet to model the IFMDB set. The DEK model is compared to the original Facenet and a transfer learning model using Triplet Loss MLP. The precision-recall curves of all models are shown below (left). The model using DEK outperforms other models.





For speaker identification, DEK is laid on top of a Convolutional Neural Network using Triplet Loss, and provides a 2% lift to the accuracy rate (right figure above).

For **supervised dimension reduction**, a trained DEK can be utilized with kernel Principal Component Analysis (kPCA). The figure below visualizes the spaces reduced using DEK and an optimized RBF kernel. The data patterns are shown clearer in the spaces mapped by DEK.



a) Cardiotocography b) Waveform c) Concrete d) Airfoil

CONCLUSIONS

DEK unifies kernel methods and deep learning into one framework with superior performance to both while mitigating their weaknesses. Experiments show that DEK outperforms typical supervised learning models in tasks like classification, regression, identity detection, and dimension reduction.

CITATION FOR FULL ARTICLES

Le, L., & Xie, Y. (2018). Deep Embedding Kernel. arXiv preprint arXiv:1804.05806.